DOCUMENT RESUME

ED 320 271 EA 021 931

AUTHOR May, John
TITLE Real World Considerations in the Development of an Effective School Incentive Program.
PUB DATE Apr 90
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (Boston, MA, April 16-20, 1990).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; Evaluation Criteria; *Incentives; *Motivation Techniques; Performance Contracts; Reinforcement; Rewards
IDENTIFIERS *South Carolina

ABSTRACT

Three criteria for an effective school incentive program that enhances motivation and improves student learning are as follows: perception of program fairness; worthwhileness of rewards; and communicability of program components. Factors that contribute to these criteria are discussed in this paper based upon the South Carolina experience. Five factors that influence perceived fairness include: (1) face validity of reward criteria; (2) technical defensibility of criteria and analytical techniques; (3) generalizability of criteria; (4) equitability of outcomes; and (5) inclusion of diagnostic indicators in the data analysis provided to the schools. Fulfillment of the second criterion, worthwhileness of rewards, can be achieved by including both monetary and nonmonetary rewards in the reward structure. Finally, clear communication of the program's purposes and criteria is necessary because understanding facilitates program acceptance. Highly consistent yearly gains should not be expected or desired by a school. If the program is motivating people to improve their performances, then the gain measure should be anticipated as volatile. From a motivational standpoint the worst scenario would be for the same schools to win and lose each year. In this situation school administrators would likely conclude that analysis and modification of instructional approaches is not worthwhile. (LMI)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED320271

Real World Considerations in the Development of an Effective School Incentive Program

John May

South Carolina Department of Education

Running Head: Real World Considerations in Developing Incentive Programs

One of the presentations in the symposium entitled
Assessing School Effectiveness:
New Directions and "Real-World" Implications
at the annual convention of the
American Educational Research Association
Boston, April 1990

EA 021 931

BEST COPY AVAILABLE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
John May
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Real World Considerations in the Development of an Effective School Incentive Program

Introduction

In order for a school incentive program to be effective, i.e., to have a motivational impact on schools which results in improved student learning, it must be perceived as fair and worthwhile, and its criteria must be understandable and readily communicated. The criteria should reflect what the local community regards as the most important indicators of performance, and the methods utilized to assess those criteria must make sense to those directly affected by their application: administrators, teachers, students, and parents. The failure to develop a program which is perceived as fair, worthwhile, and understandable will doom it to early extinction.

Based upon the South Carolina experience, this paper will attempt to outline the factors which contribute to the perceptions of fairness, worthwhileness, and communicability. The paper reflects knowledge gained over a ten year period with programs developed at both the district and state levels.

The first school incentive program in South Carolina, and one of the first in the country, was implemented in Columbia's capital city school district, Richland County School District One in 1980. This program, which included achievement, attendance, and attitudinal criteria, operated successfully for four years and served as the model for the reform legislation, passed in 1984, which established the South Carolina School Incentive Reward Program (SIRP). The SIRP is now in its fifth year of statewide implementation. The criteria for the program include student achievement gain, as determined through a matched longitudinal individual student regression analysis of test scores, student attendance, and teacher attendance. The legislation provides that no school may receive a monetary reward unless the achievement gain criterion is met. The 1989-90 statewide appropriation for the School Incentive Reward Program was $4.4 million, a figure which translated to about $29 per pupil when all three criteria were met.

Each recipient school is required to provide assurances to the South Carolina Department of Education that the reward funds will be expended in accordance with program guidelines. The school's advisory council must actively participate in the decisions relative to the use of the monies. The funds are earmarked for further improvements in the schools' programs and cannot be used to enhance the salaries of existing staff or to supplant regular school district funds.

Criterion for Success-Fairness

With respect to perceived **fairness** there are five characteristics which the program should possess. **First, the criteria for reward status should be regarded as logical and reasonable by local educators.** Face validity of the program criteria is crucial, particularly at the beginning of a new, and what some people consider, novel program. For that reason, it is very important to take advantage of the thinking of local educators in developing the program. In the formulation of the Richland One incentive program, a committee of principals, appointed by the superintendent of schools, worked under the chairmanship of the district's director of research.

At the state level, the state superintendent of education appointed a committee composed of prominent and knowledgeable educators to advise the South Carolina Department of Education in the identification of specific criteria. This group included district superintendents, principals, district office administrators, educational researchers, and representatives from higher education. A consultant, Garrett K. Mandeville, whose specialty is educational measurement and statistics, was retained to advise the South Carolina Department of Education on a continuing basis. Not only was the quality of advice

which these individuals provided important in determining logical and reasonable criteria, the credibility of the program was also enhanced by the credibility of those persons associated with its development. A survey of principals and teachers conducted during the third year of implementation indicated broad support for the program: 92% of the respondents indicated that they supported the concept of rewarding schools for achievement gains, and 85% thought that schools won rewards through goal setting and hard work. The results suggested that those affected by the program found its criteria to be reasonable.

A second characteristic which directly affects perceived fairness is the technical defensibility of the criteria and the analytical techniques utilized in the data analysis. During the first year, only, of the SIRP, a relatively sophisticated cross-sectional analysis was utilized which examined schools' achievement levels at test score points above which 75% of students scored -Q_1, above which 50% of students scored -Q_2, and above which 25% of students scored -Q_3. Using the three points was intended to provide an equitable assessment of gains, not just for the "average" student, but also for high performing and low performing students.

In each school in each grade, the percent of students scoring above each of the three points Q_1, Q_2, and Q_3 was

determined for the base school year and for the current school year. The differences between these percentages for the two years represented their gains (or changes). A Z statistic for the difference between two proportions was computed for each gain in reading and mathematics for each grade tested within each school. All the Z's for each school were then averaged to produce a single gain index for each school.

A major disadvantage of the cross-sectional gains analysis was that it compared the performance of different groups of students. Schools increasing in low scoring students due to rezoning or in-migration could be adversely affected. Conversely, schools increasing in high scoring students could benefit from this artifact. A second disadvantage was that the model adjusted for different levels of performance at only three discrete points rather than on a continuous scale. This first year analysis was not well received and was not used thereafter.

For years two through five, a matched longitudinal individual student regression analysis has been employed, and despite the complication of being somewhat difficult to communicate, has been generally well received. The availability of test data for grades 1 - 11 in South Carolina has made the matched longitudinal analysis feasible and applicable to all schools. Statewide achievement in South Carolina is measured by the Basic Skills Assessment Program

(BSAP) tests, a series of criterion referenced tests administered in six grades, and the Comprehensive Tests of Basic Skills (Form U), a norm referenced test administered in five grades. The BSAP was designed to measure reading, mathematics, and writing objectives established in 1979. The CTBS was selected from among several widely marketed achievement tests because it provided the best match to the basic skills objectives. Since the BSAP tests are not vertically equated across grades and the CTBS is not administered in contiguous grades except at 4 and 5, matched test scores cannot be directly compared and some regression method is necessary to operationalize "gain." For the remainder of this paper, therefore, gain in the South Carolina matched longitudinal analysis is understood to be the discrepancy between predicted performance, based upon prior achievement, and actual achievement.

For the last several years, longitudinal matching has been conducted for about a half-million student records annually at grade levels 1 - 11. With the exception of grade one, "pretest" data were test scores from the prior spring testing; in grade one, pretest scores were prior fall Cognitive Skills Assessment Battery data. In each case, posttest data were spring test scores. The posttests were either BSAP (grades 1,2,3,6,8) or CTBS (grades 4,5,7,9, and 11) test results.

Both non-repeaters and repeaters were included in the matching process. For non-repeaters in grades two through eleven, posttest records were matched with the pretest records from the prior grade level; for repeaters, the pretest records were from the same grade level the previous year. While there was no separate repeater analysis at grade one, no students were lost (except due to attrition) since all first graders were administered the CSAB in the fall of the year.

Posttest reading and math scale scores on the BSAP and scale scores on the CTBS were regressed onto all pretest scores. A total of forty-two such analyses were conducted, two for grade one where no repeater analysis was conducted, and four for each of the ten grades where both non-repeater and repeater analyses were possible.

In an attempt to better fit the data, square terms and a product term were included along with the linear prereading and premath terms in the regression model. To be retained a predictor had to be statistically significant ($p < .05$), based on a partial F test, and provide at least a .5% increase in explained variance. Over the years the variance explained by the equations has typically been higher for non-repeaters than for repeaters and higher in the secondary grades than in the primary grades. Multiple R^2s have ranged from about .2 in mathematics for repeaters in the primary grades to about .8 in

reading and mathematics for non-repeaters in the high school grades.

Individual student level residuals were produced by dividing the raw residuals derived from the regression analyses by their estimated standard errors. After multiplying the standardized residuals by 10, the median of these residuals or discrepancies has been determined and referred to as the School Gain Index (SGI).

Educators in South Carolina realize that the Department of Education is willing to change the methodology of computing gains if a better approach becomes available. As indicated above, the analysis was changed after year one. One of the analytical techniques currently under consideration is hierarchical linear modeling (HLM). While HLM procedures have been recommended by a number of investigators, the methodology does assume that regression lines can be fitted to individual school data and that the resulting slopes and intercepts will accurately reflect each school's performance.

Given the small numbers of matched students at each grade in many schools, it is questionable whether the slopes and intercepts can, in fact, be reliably estimated for a large number of schools. For small schools, HLM will generate a slope estimate which is heavily weighted toward the weighted average slope for all schools in the sample. The procedure,

in effect, treats small schools differently than large schools. Clearly, the procedure depends heavily upon which schools happen to be included in the sample. A contraindication for the use of HLM is the fact that about 12% of South Carolina's schools have fewer that 50 students in each grade and 8% have fewer than 20 per grade.

This discussion leads us to our third characteristic of fairness, **generalizability. For the program to be perceived as fair, the analytical techniques must be generalizable to all schools.** It is not reasonable to exclude some schools from the analysis, and it is not feasible to utilize one analytical technique for one school and a different analytical technique for another school. Some school administrators of schools failing to win an incentive award would maintain that it was the particular analytical technique used which caused them to fail to win. An advantage of the current individual student regression approach is that while utilizing all available data, its application is the same for all schools: the methodology does not change when the number of cases becomes small.

A fourth consideration in the perception of fairness is the equitability of the outcomes: which schools receive the awards. All South Carolina schools have been placed into one of five grouping categories and it is within these groups that they compete for SIRP awards. The groups are based on factors

which have been found to be related to student achievement, but which are beyond the control of the school itself. These factors include the percent of students eligible for free school lunches; the percent eligible for reduced price school lunches; teachers' educational level beyond a bachelor's degree; and the percent of students meeting the readiness standard on the Cognitive Skills Assessment Battery (CSAB). Schools in Group One have the least "resources" associated with student achievement, while schools in Group Five have the most "resources."

In the most recent grouping analyses, conducted during the 1988-89 school year, more than 40 context indicators were examined for their potential to predict student achievement and thus be utilized as grouping variables. These indicators can be classified into one of three domains: (1) student home characteristics, (2) community characteristics, and (3) school resources. Data were collected on each of these predictors and then aggregated to the school or district level. Stepwise multiple regressions were computed for each of the various grades and subtests for three years of data. The variables most predictive of achievement for each of the subtests (reading and mathematics) and grades (1-11) were identified and their regression coefficients and correlations examined for consistency. The four variables described above most frequently accounted for as much variance as could be accounted for in relationship to achievement.

For the past four years SIRP awards have gone to the schools with SGIs in the top quarter of the schools in each grouping category. It has been emphasized that while the SGI and grouping category are correlated (about .4), the SGI and recipient status are uncorrelated. During this time there has been little, if any, criticism of this approach. This acceptance comes in spite of the fact that some non-winning schools in higher grouping categories have higher school gain indices than winning schools in lower grouping categories. There is apparently an understanding that equity and perceived fairness are closely tied.

The fifth characteristic of an incentive program which is perceived as fair is one in which diagnostic indicators are included in the data analysis provided to the schools. As noted above, school administrators want feedback in order to determine where to place priorities and where improvements need to be made. If they are not provided information which can be used to help target instruction, many will reject the program as unfair. It is for this reason that individual student residuals are aggregated at the grade and subtest level and presented in South Carolina's School Performance Reports system. Using this report principals can identify which grades or skills negatively impacted their overall school gain index and can direct their energies toward improved performance in those grades or subtests.

A variety of interesting patterns of gain index performance has emerged from analyses of the data. Some schools and school districts, for example, show low gains in reading and relatively higher gains in mathematics while in other schools and districts the reverse is the case. Some schools have gains which are high in the grades tested with the South Carolina's criterion referenced tests (BSAP) and low in the grades in which the norm referenced tests are administered. Other schools show higher performance in the norm referenced grades and lower performance in the BSAP tested grades. The implication is that in many cases schools' and districts' curricula are focused toward certain objectives to the exclusion of others. These analyses have prompted schools and districts to examine the articulation, coordination, and alignment of their curricula.

In addition to the gains analysis, the School Performance Report presents comparative data on level of achievement, attendance, and dropout data for a three year period. Comparisons of performance with all other schools in the state and with other similar schools are provided, making this report a powerful diagnostic tool. Schools can determine where achievement levels need improvement, where achievement gains are low, and how the schools are doing with respect to student attendance, teacher attendance, and student dropout rates.

Criterion for Success-Worthwhileness of Rewards

A second major factor in the success of a school incentive program is the perceived worthwhileness of the rewards. As indicated above, the South Carolina program provides both monetary and non-monetary incentives to schools which achieve gains that place them in the top quarter of their school grouping category. Over the last several years the amount of the monetary reward has ranged from about $25 to $30 per pupil in enrollment. Thus, for the average school with a student population of 500, the award amount has exceeded $12,000. While the relationship between the per pupil reward amount and motivation has not been investigated systematically, anecdotal accounts indicate that this per pupil amount has been sufficient to spur motivation.

The SIRP also provides non-monetary reward objects including flags and certificates to schools as symbols of their accomplishment. Non-monetary rewards have been effectively utilized by a variety of institutions and organizations from the military to the boy scouts. For some school administrators and teachers, the objects which symbolize their program's success appear to be as important as the monetary rewards themselves. Some schools fly their flags on their flag pole while others hang them in a prominent place

such as the school's foyer, cafeteria, or gymnasium. One principal intimated that he intended to hang his school's flags in the auditorium just like the Celtics hang their's in the Boston Garden.

Reward certificates are provided to both monetary award winners and to honorable mention recipients. Honorable mention schools are those which approach but do not quite meet the criteria for monetary recipient status. These schools receive engraved silver foil certificates signifying their accomplishments. Monetary award winners, on the other hand, get gold foil certificates. In order to make the program more meaningful to students, and to reinforce the improvement impulse, the Department of Education has recently begun providing certificates which recognize individual students who have demonstrated exceptional improvement in achievement. These awards are presented to two students per grade level in each school in the state.

The provision of certificates for honorable mention recipients is designed to expand the number of schools positively affected by the program. By including a sizable number of schools in the reward structure, it is hoped that more schools will view winner status as attainable. If winner status is seen as unattainable, then the worthwhileness of the effort required to reach the status will be diminished.

The perceived worthwhileness of a program is enhanced by the fact that it is virtually paperwork free. There is no requirement for application, no grant writing, and no data analysis required at the school level. All data are collected and analyzed by the State Department of Education. Schools which are School Incentive Reward Program recipients are, however, required to submit an assurance form stipulating the schools' compliance with program requirements.

Criterion for Success-Communicability

A final consideration in the successful implementation of a school incentive program is the clear communication of the program's purposes and criteria. Separate descriptions of the program for lay, educator, and technical audiences should be developed and sufficient staff time must be set aside for communications efforts, a substantial portion of which should precede the announcement of rewards. Most audiences are far more willing to listen to and accept the rationale and criteria for an awards program prior to the announcement of recipients.

In South Carolina the major communications problem has been the articulation of what the gain index represents. Despite numerous workshops, inservice activities, brochures, video tapes and hundred of conversations with school

administrators, the interpretation of the gain index remains a critically important effort. After many attempts at refinement, the current approach utilizes examples and analogies to communicate the nature of the analysis.

The communication of the gain index is complicated by the prior experience of school administrators with "gains" in student achievement. The traditional cross-sectional approach to achievement gains, i.e., comparing this year's third graders with last year's third graders, or this year's third graders with last year's second graders, actually interferes with the understanding of matched longitudinal gains based upon regression analysis. Older ways of thinking have to be discarded as well as some convenient crutches. For years educators have explained that while this year's third grade scores are somewhat lower that last year's third grade scores, these are, after all, different groups of students. Even when looking at the same cohort of students, these are not in most cases identical groups of students. Some second graders were not promoted to the third grade and there has been student migration between the schools over the course of the year. Traditionally, when educators have been hard pressed to explain drops in student achievement, they have sometimes reasoned that the factors of student mobility, student attrition, and other changes in the school's population have been, at least in part, responsible for the changes in achievement.

The matched longitudinal individual regression analysis narrows considerably the "wiggle room" school administrators have enjoyed. While factors other than instruction may certainly influence a school's gain index, many of the standard explanations for relatively lower performance are no longer viable. As school principals come face to face with the data suggesting that their students have not made gains comparable to those made by students in other similar schools in South Carolina, their first reaction is to question the data and the methodology which led to such a dastardly outcome. It is in these circumstances that clarity of communication is most acutely tested. The interpersonal and interpretive skills required to communicate a highly technical concept to a local school administrator who feels threatened or even insulted by the data should not in any way be underestimated. Bridging the information gap and bringing the local administrator to the point of examining ways of improving his school's educational program is, perhaps, the most difficult challenge for the administrator of a school incentive program.

The use of the HLM methodology discussed earlier would do nothing to simplify communications since it is technically more complex than individual student regression. It would be necessary, for example, to explain to lay persons the differences between within and between-school effects, and slopes and intercepts. Between-school effects would pose a

particularly difficult explicative problem. Earlier it was noted that award decisions for the SIRP are currently made within five comparison groups reflecting differences in school context. While the model utilized to identify school grouping categories may well be imperfect, it is explicit. School administrators understand the variables used and how schools are placed in a particular category. It is clear that each school's gain is compared with those of other "similar" schools. With HLM, on the other hand, school context differences are addressed in the between-schools model. The context data would enter directly into the calculational procedure, making the gain index derivation even more obscure than it is currently.

A Final Issue-Stability of Gains and School Effectiveness

A concern often raised in the literature is the stability and consistency of school "effectiveness." It is often maintained that a key element in the success of a school incentive program is the stability of gains. The South Carolina reform act, which required the Department of Education to identify criteria, including achievement gain to recognize and reward schools, made no explicit reference to "effective schools". The intent of the law was to motivate school faculties to work toward improved performance. The

South Carolina Department of Education has adopted a conservative approach to the issue of school effectiveness: no claim has been made that the gains analysis identifies the most "effective" schools. The reasoning here has been that schools may vary from year to year in the degree of gain manifested and that this, in and of itself, is not indicative of effectiveness. The Department of Education's position on the issue of gains and effectiveness avoids the problem of labeling schools that do not win incentive awards as "ineffective". After five years of operation of the program, it has become clear that "low gain" is a stigma: the tacit identification of a school as "ineffective" or "less effective" would exacerbate the problem of labeling and make the successful operation of the program more difficult.

Given the nature of gain, it should neither be expected nor desired that the gains shown by a school be highly consistent from one year to the next. If the program is having its desired effect, i.e., to motivate school administrators, teachers, parents and students to improve performance, then it should be anticipated that the gain measure will be rather volatile. In a survey of school principals and teachers, 85% reported that winning a School Incentive Reward was an important school goal and more than two-thirds indicated that the School Incentive Program provided additional motivation to the schools. The program has inspired school staffs to carefully assess their

instructional strategies and to modify them in such a way as to improve student learning.

During the first year of the program's implementation, nearly all of the inservice sessions provided to the schools by the incentive program's staff focused on the program's criteria, the calculation of the School Gain Index, and the appropriate expenditure of the funds. Over time, the emphasis in these workshops has shifted to strategies which a school might implement to win the awards. Sharing information on what techniques have been reported to be successful in winning the award have come to dominate these discussions. Hardly a week goes by that a school principal does not call the Department of Education for references on which schools have been successful in attaining incentive recipient status.

The percentage of schools repeating as winners from one year to the next has been about 40%. From a motivational standpoint the worst thing that could happen would be for the same schools to win and lose each year. If this were to occur, school administrators would likely conclude that there is no point in putting forth the effort to examine strategies and approaches to instruction since the same schools are going to win or lose each year. On the other hand, if the percentage of repeaters was very low, the conclusion could be drawn that the gains statistic is like a roller coaster: if you do well one year, you are likely to do poorly the next

year. Again, such a result would not support putting forth the effort to examine programs and curriculum.

Table 1 below presents correlations of school level gains by grade and subtest and also by total school.

Table 1

Correlations (Stability) of School Level Gains by Grade and Total School Gain Index (SGI)

Grade	1988 to 1989 Reading	1988 to 1989 Math	1987 to 1989 Reading	1987 to 1989 Math
1	.588	.530	.429	.330
2	.484	.396	.468	.352
3	.343	.426	.276	.329
4	.553	.531	.461	.395
5	.380	.386	.441	.391
6	.424	.451	.422	.425
7	.531	.480	.563	.415
8	.590	.476	.288	.150
9	.725	.561	.683	.467
10	.436	.520	.438	.388
11	.603	.497	.617	.527
Total SGI	.433		.484	

Inspection of table 1 reveals that there is a modest relationship of about .4 between a school's gain index in one year and its school gain index in the next. Grade subtest gain indices range from about .3 to .7.

Summary and Conclusions

This paper has suggested three criteria for an effective school incentive program, i.e, one which has a motivational impact upon schools and results in improved student learning: (a) it must be perceived as fair; (b) the awards and recognition must be considered worthwhile; and (c) its components must be understandable and readily communicated.

With respect to perceived fairness there are five characteristics which the program should possess. First, the criteria for reward status should have face validity and be regarded as logical and reasonable by local educators. The inclusion of knowledgeable and respected educators in the planning and implementation of the program is critical to its success.

A second characteristic which directly affects perceived fairness is the technical defensibility of the criteria and

the analytical techniques utilized in the data analysis. The advantages of matched longitudinal over cross-sectional analyses were discussed. The utilization of HLM methodologies versus individual student regression techniques were addressed. Potential limitations in applicability of HLM were enumerated: these included the estimation of slopes and intercepts for individual schools when the number of cases available for the analysis is small.

The third characteristic of fairness is generalizability: it is not reasonable to exclude some schools from the analysis because of problems such as small sample sizes, and it is not feasible to utilize one analytical technique for one school and a different analytical technique for another school.

A fourth consideration in the perception of fairness is the equitability of the outcomes. All schools, regardless of the background characteristics of the students and school community, should have a fair opportunity to win the rewards and recognition afforded by the program. In South Carolina, school socioeconomic context is controlled by having schools compete for awards within school grouping categories.

The fifth characteristic of an incentive program which is perceived as fair is one in which diagnostic indicators are included in the data analysis provided to the schools. School administrators need information from the analyses that can be

used to help target instruction. In the absence of such data, many will consider the program to be unfair because they do not know what areas need improvement.

In addition to fairness, a second major factor in the success of a school incentive program is the perceived worthwhileness of the rewards. It was suggested that both monetary and non-monetary rewards be included in the reward structure.

The final factor necessary for the successful implementation of a school incentive program is the clear communication of the program's purposes and criteria. If school administrators cannot understand the criteria by which the schools are judged and communicate them to students, teachers, and parents, they will be inclined to reject to program.

Given the nature of gain, it should neither be expected nor desired that the gains shown by a school be highly consistent from one year to the next. If the program is having its desired effect, i.e., to motivate school administrators, teachers, parents and students to improve performance, then it should be anticipated that the gain measure will be rather volatile. From a motivational standpoint the worst thing that could happen would be for the same schools to win and lose each year. If this were to

occur, school administrators would likely conclude that there is no point in putting forth the effort to examine and modify strategies and approaches to instruction.